# SPOTIFY ATTRIBUTES REGRESSION PROJECT

Kaustubh Dangche, Lakshmi Sindhu Pulugundla & Lasya Priya Thota

2024-11-08

## Project Objective and Overview

The objective of this project is to predict the `msPlayed` (playtime in milliseconds) of songs on **Spotify** based on various song attributes. These attributes include both **continuous** variables such as **danceability**, **energy**, **loudness**, and **tempo**, as well as **categorical** features like **genre** and **artistName**.

The project's core goal is to understand how different musical attributes influence the total playtime of songs and build a regression model to predict future playtime. The dataset includes a variety of song characteristics that provide insights into musical features, which can be useful for music industry professionals and data scientists alike.

By the end of this project, the following objectives will be achieved: - Build a robust **regression model** to predict `msPlayed`. - Understand the significance of various musical features in predicting playtime. - Clean and preprocess the data, handling missing values, duplicates, and categorical features. - Perform **Exploratory Data Analysis (EDA)** to visualize trends and patterns in the data.

---

**Focus of the Project:**

- **Primary Application**: **Music Playtime Prediction**
  - Predicting the playtime (`msPlayed`) based on the attributes of the song.
- **Core Model**: **Regression Model (Multiple Linear Regression / Polynomial Regression)**
  - Using regression analysis to model the relationship between song attributes and playtime.

---

**Phase-Wise Project Flow**

**Phase 1: Data Preprocessing & Basic Data Understanding**

1. **Data Loading and Initial Exploration**
   - Loading the Spotify Song Attributes dataset and inspecting its structure.
2. **Data Cleaning and Type Modification**
   - Converting columns into appropriate formats and removing redundant or irrelevant columns.
3. **Handling Missing Values and Checking for Duplicates**

- Ensuring no missing values or duplicate records exist.

4. **Outlier Detection and Handling**
   - Identifying and addressing any extreme outliers in key features.
5. **Feature Scaling**
   - Standardizing and normalizing continuous features, making them suitable for regression modeling.

---

### Phase 2: Exploratory Data Analysis (EDA)

1. **Visualizing Distribution of Attributes**
   - Visualizing how each feature, especially continuous variables like `danceability`, `energy`, and `msPlayed`, is distributed.
2. **Exploring Relationships Between Features**
   - Using scatter plots, correlation matrices, and other visual tools to explore relationships between the features and `msPlayed`.

---

### Phase 3: Feature Engineering & Extraction

1. **Handling Categorical Data**
   - Converting categorical variables like `genre` and `artistName` into factors, and preparing them for model fitting.
2. **Creating New Features**
   - Deriving new features from the existing ones to enhance the predictive power of the model.

---

### Phase 4: Model Building

1. **Model Selection (Linear/Multiple Regression)**
   - Selecting a suitable regression model to predict `msPlayed`.
2. **Model Evaluation**
   - Evaluating the model performance using metrics such as R-squared, RMSE, and residual plots.

---

### Phase 5: Post-Modeling Analysis

1. **Model Refinement**
   - Making improvements to the model based on diagnostic tests and evaluation metrics.

---

**Phase 6: Conclusion and Future Enhancements**

1. **Summarizing Findings**
   - Discussing the effectiveness of the regression model and the insights gained.
2. **Future Work**
   - Suggesting improvements, such as trying more complex models (e.g., Random Forest, Gradient Boosting), or exploring additional features for better predictions.

---

# 1. Data Collection and Pre-Processing

## 1.1 Load the Necessary Libraries & Dataset

```
# Load necessary libraries
library(dplyr)  # For data manipulation
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)  # Optional, for visualization if needed

# Load the Spotify attribute dataset
data <- read.csv("Spotify_Song_Attributes.csv")
# Create a copy of the spotify attribute dataset
spotify_data <- data
# View the first few rows to understand the structure
head(spotify_data)
```

```
##                                                              trackName
## 1                                                             "Honest"
## 2 "In The Hall Of The Mountain King" from Peer Gynt Suite N°1, Op. 46
## 3                                                   #BrooklynBloodPop!
## 4                                                                  $10
## 5                                              (I Just) Died In Your Arms
## 6                                                         (L)only Child
##                 artistName msPlayed             genre danceability energy key
## 1             Nico Collins   191772                          0.476  0.799   4
## 2 London Symphony Orchestra  1806234 british orchestra        0.475  0.130   7
## 3                     SyKo   145610        glitchcore         0.691  0.814   1
```

```
## 4              Good Morning    25058  experimental pop        0.624  0.596   4
## 5              Cutting Crew  5504949        album rock        0.625  0.726  11
## 6               salem ilese  2237969             alt z        0.645  0.611   8
##   loudness mode speechiness acousticness instrumentalness liveness valence
## 1   -4.939    0      0.2120       0.0162         0.00e+00   0.2570   0.577
## 2  -17.719    1      0.0510       0.9160         9.56e-01   0.1010   0.122
## 3   -3.788    0      0.1170       0.0164         0.00e+00   0.3660   0.509
## 4   -9.804    1      0.0314       0.4750         2.03e-01   0.1190   0.896
## 5  -11.402    0      0.0444       0.0158         1.69e-04   0.0625   0.507
## 6   -5.925    0      0.1370       0.2900         2.05e-05   0.2370   0.645
##     tempo           type                     id
## 1 162.139 audio_features 7dTxqsaFGHOXwtzHINjfHv
## 2 112.241 audio_features 14Qcrx6DfjvcjOH8oV8oUW
## 3 132.012 audio_features 7K9Z3yFNNLv5kwTjQYGjnu
## 4 120.969 audio_features 3koAwrM1RO0TGMeQJ3qt9J
## 5 124.945 audio_features 4ByEFOBuLXpCqvO1kw8Wdm
## 6 157.475 audio_features 22lJaG2yxlSjIwdUIddcFk
##                                    uri
## 1 spotify:track:7dTxqsaFGHOXwtzHINjfHv
## 2 spotify:track:14Qcrx6DfjvcjOH8oV8oUW
## 3 spotify:track:7K9Z3yFNNLv5kwTjQYGjnu
## 4 spotify:track:3koAwrM1RO0TGMeQJ3qt9J
## 5 spotify:track:4ByEFOBuLXpCqvO1kw8Wdm
## 6 spotify:track:22lJaG2yxlSjIwdUIddcFk
##                                                 track_href
## 1 https://api.spotify.com/v1/tracks/7dTxqsaFGHOXwtzHINjfHv
## 2 https://api.spotify.com/v1/tracks/14Qcrx6DfjvcjOH8oV8oUW
## 3 https://api.spotify.com/v1/tracks/7K9Z3yFNNLv5kwTjQYGjnu
## 4 https://api.spotify.com/v1/tracks/3koAwrM1RO0TGMeQJ3qt9J
## 5 https://api.spotify.com/v1/tracks/4ByEFOBuLXpCqvO1kw8Wdm
## 6 https://api.spotify.com/v1/tracks/22lJaG2yxlSjIwdUIddcFk
##                                                         analysis_url duration_ms
## 1 https://api.spotify.com/v1/audio-analysis/7dTxqsaFGHOXwtzHINjfHv      191948
## 2 https://api.spotify.com/v1/audio-analysis/14Qcrx6DfjvcjOH8oV8oUW      150827
## 3 https://api.spotify.com/v1/audio-analysis/7K9Z3yFNNLv5kwTjQYGjnu      145611
## 4 https://api.spotify.com/v1/audio-analysis/3koAwrM1RO0TGMeQJ3qt9J       89509
## 5 https://api.spotify.com/v1/audio-analysis/4ByEFOBuLXpCqvO1kw8Wdm      280400
## 6 https://api.spotify.com/v1/audio-analysis/22lJaG2yxlSjIwdUIddcFk      144468
##   time_signature
## 1              4
## 2              4
## 3              4
## 4              4
## 5              4
## 6              3
```

**Inference:**

The dataset is successfully loaded into R, and a copy (`spotify_data`) is created to preserve the original data. We can now inspect the first few rows to understand its structure.

## 1.2 View Dataset Structure

```
# View the structure of the data
str(spotify_data)
```

```
## 'data.frame':    10080 obs. of  22 variables:
##  $ trackName       : chr  "\"Honest\"" "\"In The Hall Of The Mountain King\" from Peer Gynt Suite N°
##  $ artistName      : chr  "Nico Collins" "London Symphony Orchestra" "SyKo" "Good Morning" ...
##  $ msPlayed        : int  191772 1806234 145610 25058 5504949 2237969 441335 70589 120005 107407 ...
##  $ genre           : chr  "" "british orchestra" "glitchcore" "experimental pop" ...
##  $ danceability    : num  0.476 0.475 0.691 0.624 0.625 0.645 0.663 NA 0.792 0.759 ...
##  $ energy          : num  0.799 0.13 0.814 0.596 0.726 0.611 0.904 NA 0.511 0.699 ...
##  $ key             : num  4 7 1 4 11 8 7 NA 2 0 ...
##  $ loudness        : num  -4.94 -17.72 -3.79 -9.8 -11.4 ...
##  $ mode            : num  0 1 0 1 0 0 1 NA 1 0 ...
##  $ speechiness     : num  0.212 0.051 0.117 0.0314 0.0444 0.137 0.0857 NA 0.0409 0.0307 ...
##  $ acousticness    : num  0.0162 0.916 0.0164 0.475 0.0158 0.29 0.000708 NA 0.124 0.202 ...
##  $ instrumentalness: num  0.00 9.56e-01 0.00 2.03e-01 1.69e-04 2.05e-05 2.89e-01 NA 9.04e-05 1.31e-04
##  $ liveness        : num  0.257 0.101 0.366 0.119 0.0625 0.237 0.341 NA 0.14 0.443 ...
##  $ valence         : num  0.577 0.122 0.509 0.896 0.507 0.645 0.675 NA 0.111 0.907 ...
##  $ tempo           : num  162 112 132 121 125 ...
##  $ type            : chr  "audio_features" "audio_features" "audio_features" "audio_features" ...
##  $ id              : chr  "7dTxqsaFGHOXwtzHINjfHv" "14Qcrx6DfjvcjOH8oV8oUW" "7K9Z3yFNNLv5kwTjQYGjnu"
##  $ uri             : chr  "spotify:track:7dTxqsaFGHOXwtzHINjfHv" "spotify:track:14Qcrx6DfjvcjOH8oV8oU
##  $ track_href      : chr  "https://api.spotify.com/v1/tracks/7dTxqsaFGHOXwtzHINjfHv" "https://api.sp
##  $ analysis_url    : chr  "https://api.spotify.com/v1/audio-analysis/7dTxqsaFGHOXwtzHINjfHv" "https:/
##  $ duration_ms     : num  191948 150827 145611 89509 280400 ...
##  $ time_signature  : num  4 4 4 4 4 3 4 NA 4 4 ...
```

**Inference:**

The `str()` function provides a quick look at the internal structure of the dataset. We can check the data types (e.g., numeric, factor) of each column, which helps in deciding how to clean or transform the data.

## 1.3 Summary of Each Column

```
# View a summary of each column (e.g., min, max, mean, etc.)
summary(spotify_data)
```

```
##   trackName          artistName           msPlayed             genre
##  Length:10080       Length:10080       Min.   :        0   Length:10080
##  Class :character   Class :character   1st Qu.:   136780   Class :character
##  Mode  :character   Mode  :character   Median :   266288   Mode  :character
##                                        Mean   :  1519657
##                                        3rd Qu.:  1186307
##                                        Max.   :158367130
##
##   danceability        energy             key            loudness
##  Min.   :0.0000   Min.   :0.0011   Min.   : 0.000   Min.   :-42.044
```

```
##  1st Qu.:0.5090   1st Qu.:0.4030   1st Qu.: 2.000   1st Qu.:-10.189
##  Median :0.6230   Median :0.5890   Median : 5.000   Median : -7.218
##  Mean   :0.6025   Mean   :0.5635   Mean   : 5.242   Mean   : -8.685
##  3rd Qu.:0.7140   3rd Qu.:0.7510   3rd Qu.: 8.000   3rd Qu.: -5.336
##  Max.   :0.9760   Max.   :0.9990   Max.   :11.000   Max.   :  3.010
##  NA's   :550      NA's   :550      NA's   :550      NA's   :550
##       mode         speechiness      acousticness    instrumentalness
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0361   1st Qu.:0.0538   1st Qu.:0.0000
##  Median :1.0000   Median :0.0479   Median :0.2450   Median :0.0000
##  Mean   :0.6124   Mean   :0.0785   Mean   :0.3629   Mean   :0.1532
##  3rd Qu.:1.0000   3rd Qu.:0.0819   3rd Qu.:0.6680   3rd Qu.:0.0276
##  Max.   :1.0000   Max.   :0.9660   Max.   :0.9960   Max.   :0.9930
##  NA's   :550      NA's   :550      NA's   :550      NA's   :550
##     liveness         valence           tempo            type
##  Min.   :0.0249   Min.   :0.0000   Min.   :  0.00   Length:10080
##  1st Qu.:0.0962   1st Qu.:0.2370   1st Qu.: 97.57   Class :character
##  Median :0.1190   Median :0.4090   Median :119.82   Mode  :character
##  Mean   :0.1746   Mean   :0.4341   Mean   :119.37
##  3rd Qu.:0.2090   3rd Qu.:0.6140   3rd Qu.:139.78
##  Max.   :0.9640   Max.   :0.9860   Max.   :236.20
##  NA's   :550      NA's   :550      NA's   :550
##       id                uri             track_href        analysis_url
##  Length:10080       Length:10080       Length:10080       Length:10080
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##   duration_ms      time_signature
##  Min.   :  10027   Min.   :0.000
##  1st Qu.: 161697   1st Qu.:4.000
##  Median : 194286   Median :4.000
##  Mean   : 202931   Mean   :3.917
##  3rd Qu.: 229526   3rd Qu.:4.000
##  Max.   :4581483   Max.   :5.000
##  NA's   :550       NA's   :550
```

**Inference:**

The summary statistics allow us to understand the distribution of each feature, including mean, min, max, and standard deviation. This is crucial for identifying outliers and deciding which variables require special treatment.

## 1.4 Handling Missing Values

```r
# Check for missing values in each column
colSums(is.na(spotify_data))
```

```
##        trackName       artistName         msPlayed            genre
```

```
##                 0                 0                 0                 0
##      danceability            energy               key          loudness
##               550               550               550               550
##              mode       speechiness      acousticness  instrumentalness
##               550               550               550               550
##          liveness           valence             tempo              type
##               550               550               550                 0
##                id               uri        track_href      analysis_url
##                 0                 0                 0                 0
##       duration_ms    time_signature
##               550               550
```

**We can see:**

1. **Columns with Zero Missing Values**: Columns like `trackName`, `artistName`, `msPlayed`, `genre`, `id`, `uri`, `track_href`, `analysis_url`, and `type` have no missing values, ensuring they are complete for analysis.
2. **Columns with 550 Missing Values**: Columns such as `danceability`, `energy`, `key`, `loudness`, `mode`, `speechiness`, `acousticness`, `instrumentalness`, `liveness`, `valence`, `tempo`, `duration_ms`, and `time_signature` have significant missing values (550). This indicates a need to address these missing entries to ensure reliable analysis.

### 1.4.1 Remove Missing Values

```r
# Option 1: Remove rows with missing values (if they are few)
spotify_data <- na.omit(spotify_data)
```

**Inference:**

Using `na.omit()` removes rows with missing data, leaving us with only complete observations. This is necessary for regression models, which cannot handle missing values.

### 1.4.2 Check Missing Values After Removal

```r
# Check for missing values in each column after applying na.omit function
colSums(is.na(spotify_data))
```

```
##         trackName        artistName          msPlayed             genre
##                 0                 0                 0                 0
##      danceability            energy               key          loudness
##                 0                 0                 0                 0
##              mode       speechiness      acousticness  instrumentalness
##                 0                 0                 0                 0
##          liveness           valence             tempo              type
##                 0                 0                 0                 0
##                id               uri        track_href      analysis_url
##                 0                 0                 0                 0
##       duration_ms    time_signature
##                 0                 0
```

**All values being zero in the output confirms that no missing data remains in the dataset. This clean dataset now contains only complete rows, ideal for accurate and consistent analysis.**

**Just check missing values removed from structure data or not.**

```
str(spotify_data)
```

```
## 'data.frame':    9530 obs. of  22 variables:
##  $ trackName       : chr  "\"Honest\"" "\"In The Hall Of The Mountain King\" from Peer Gynt Suite N°
##  $ artistName      : chr  "Nico Collins" "London Symphony Orchestra" "SyKo" "Good Morning" ...
##  $ msPlayed        : int  191772 1806234 145610 25058 5504949 2237969 441335 120005 107407 21354 ...
##  $ genre           : chr  "" "british orchestra" "glitchcore" "experimental pop" ...
##  $ danceability    : num  0.476 0.475 0.691 0.624 0.625 0.645 0.663 0.792 0.759 0.742 ...
##  $ energy          : num  0.799 0.13 0.814 0.596 0.726 0.611 0.904 0.511 0.699 0.675 ...
##  $ key             : num  4 7 1 4 11 8 7 2 0 8 ...
##  $ loudness        : num  -4.94 -17.72 -3.79 -9.8 -11.4 ...
##  $ mode            : num  0 1 0 1 0 0 1 1 0 1 ...
##  $ speechiness     : num  0.212 0.051 0.117 0.0314 0.0444 0.137 0.0857 0.0409 0.0307 0.0907 ...
##  $ acousticness    : num  0.0162 0.916 0.0164 0.475 0.0158 0.29 0.000708 0.124 0.202 0.139 ...
##  $ instrumentalness: num  0.00 9.56e-01 0.00 2.03e-01 1.69e-04 2.05e-05 2.89e-01 9.04e-05 1.31e-04 0
##  $ liveness        : num  0.257 0.101 0.366 0.119 0.0625 0.237 0.341 0.14 0.443 0.167 ...
##  $ valence         : num  0.577 0.122 0.509 0.896 0.507 0.645 0.675 0.111 0.907 0.679 ...
##  $ tempo           : num  162 112 132 121 125 ...
##  $ type            : chr  "audio_features" "audio_features" "audio_features" "audio_features" ...
##  $ id              : chr  "7dTxqsaFGHOXwtzHINjfHv" "14Qcrx6DfjvcjOH8oV8oUW" "7K9Z3yFNNLv5kwTjQYGjnu"
##  $ uri             : chr  "spotify:track:7dTxqsaFGHOXwtzHINjfHv" "spotify:track:14Qcrx6DfjvcjOH8oV8o
##  $ track_href      : chr  "https://api.spotify.com/v1/tracks/7dTxqsaFGHOXwtzHINjfHv" "https://api.sp
##  $ analysis_url    : chr  "https://api.spotify.com/v1/audio-analysis/7dTxqsaFGHOXwtzHINjfHv" "https:/
##  $ duration_ms     : num  191948 150827 145611 89509 280400 ...
##  $ time_signature  : num  4 4 4 4 4 3 4 4 4 4 ...
##  - attr(*, "na.action")= 'omit' Named int [1:550] 8 48 77 78 109 123 125 129 132 142 ...
##   ..- attr(*, "names")= chr [1:550] "8" "48" "77" "78" ...
```

## 1.5 Remove Unwanted Columns

```
# Load dplyr library
library(dplyr)

# Remove `trackName` along with other unnecessary columns
spotify_data <- spotify_data %>% select(-id, -uri, -track_href, -analysis_url, -trackName, -type)
```

**Statistical Reasoning for Dropping Variables:**

- **`id`, `uri`, `track_href`, `analysis_url`, `trackName`, `type`**: These columns do not contribute to predicting `msPlayed`. They are either unique identifiers or non-informative for the analysis.

```
# Check the specified columns have been removed or not
str(spotify_data)
```

```
## 'data.frame':    9530 obs. of  16 variables:
##  $ artistName      : chr  "Nico Collins" "London Symphony Orchestra" "SyKo" "Good Morning" ...
##  $ msPlayed        : int  191772 1806234 145610 25058 5504949 2237969 441335 120005 107407 21354 ...
##  $ genre           : chr  "" "british orchestra" "glitchcore" "experimental pop" ...
##  $ danceability    : num  0.476 0.475 0.691 0.624 0.625 0.645 0.663 0.792 0.759 0.742 ...
##  $ energy          : num  0.799 0.13 0.814 0.596 0.726 0.611 0.904 0.511 0.699 0.675 ...
##  $ key             : num  4 7 1 4 11 8 7 2 0 8 ...
##  $ loudness        : num  -4.94 -17.72 -3.79 -9.8 -11.4 ...
##  $ mode            : num  0 1 0 1 0 0 1 1 0 1 ...
##  $ speechiness     : num  0.212 0.051 0.117 0.0314 0.0444 0.137 0.0857 0.0409 0.0307 0.0907 ...
##  $ acousticness    : num  0.0162 0.916 0.0164 0.475 0.0158 0.29 0.000708 0.124 0.202 0.139 ...
##  $ instrumentalness: num  0.00 9.56e-01 0.00 2.03e-01 1.69e-04 2.05e-05 2.89e-01 9.04e-05 1.31e-04 0
##  $ liveness        : num  0.257 0.101 0.366 0.119 0.0625 0.237 0.341 0.14 0.443 0.167 ...
##  $ valence         : num  0.577 0.122 0.509 0.896 0.507 0.645 0.675 0.111 0.907 0.679 ...
##  $ tempo           : num  162 112 132 121 125 ...
##  $ duration_ms     : num  191948 150827 145611 89509 280400 ...
##  $ time_signature  : num  4 4 4 4 4 3 4 4 4 4 ...
##  - attr(*, "na.action")= 'omit' Named int [1:550] 8 48 77 78 109 123 125 129 132 142 ...
##   ..- attr(*, "names")= chr [1:550] "8" "48" "77" "78" ...
```

## 1.6 Data Cleaning

### 1.6.1 Check Whitespace in artistName Column

```r
# Check for leading or trailing whitespace
whitespace_issues <- spotify_data$artistName[grepl("^\\s|\\s$", spotify_data$artistName)]

# Display entries with whitespace issues
print(whitespace_issues)
```

```
## character(0)
```

**Inference:**

Whitespace in `artistName` can cause inconsistencies, so it is necessary to clean these entries.

### 1.6.2 Identify Special Characters in artistName

```r
# Identify special characters in artist names
special_characters <- spotify_data$artistName[grepl("[^a-zA-Z0-9\\s]", spotify_data$artistName)]

# Display total number of entries with special characters and the first 25 entries
cat("Total number of ntries with Special Characters:", length(special_characters), "\n")
```

```
## Total number of ntries with Special Characters: 6030
```

```r
head(special_characters, 25)
```

```
##  [1] "Nico Collins"                "London Symphony Orchestra"
##  [3] "Good Morning"                "Cutting Crew"
##  [5] "salem ilese"                 "Eren Cannata"
##  [7] "$uicideboy$"                 "Britney Spears"
##  [9] "Sidhu Moose Wala"            "colours in the dark"
## [11] "Dan + Shay"                  "mj apanay"
## [13] "Chris James"                 "Zach Hood"
## [15] "Josh Golden"                 "Shiloh Dynasty"
## [17] "Jeremy Zucker"               "Thomas Newman"
## [19] "Bo Burnham"                  "Linkin Park"
## [21] "Marco Luka"                  "Sleepy Hallow"
## [23] "Comfort Club"                "Johannes Brahms"
## [25] "Seattle Symphony Orchestra"
```

**Inference:**

Special characters in `artistName` can cause inconsistencies when grouping artists. Identifying these characters helps ensure that they are cleaned properly.

### 1.6.3 Identifying Duplicated Artist Names

```r
# Identify duplicated artist names after cleaning
duplicate_artists <- spotify_data$artistName[duplicated(spotify_data$artistName)]

# Display total number of duplicates and the first 10 duplicate entries
cat("Total number of duplicates:", length(duplicate_artists), "\n")
```

```
## Total number of duplicates: 7343
```

```r
head(duplicate_artists, 10)
```

```
##  [1] "glaive"                   "Johannes Brahms"
##  [3] "Lauv"                     "Chris James"
##  [5] "Ed Sheeran"               "Ed Sheeran"
##  [7] "blackbear"                "blackbear"
##  [9] "High School Musical Cast" "KUTE"
```

**Inference:**

This output shows the total number of duplicate artist names and the first 10 duplicates, helping confirm that duplicate names are properly handled.

### 1.6.4 Replace Special Characters

```r
# Replace special characters like $ with s (only if necessary)
spotify_data$artistName <- gsub("\\$", "s", spotify_data$artistName)
```

**Inference:**

Replacing special characters ensures that `artistName` values are consistent, removing unwanted characters that may interfere with analysis.

#### 1.6.5 Convert Artist Names to Lowercase

```
# Convert artistName to lowercase
spotify_data$artistName <- tolower(spotify_data$artistName)
```

**Inference:**

Lowercasing ensures that variations in capitalization are removed, making artist names uniform for analysis.

#### 1.6.6 Removing Identical Rows

```
# Removing rows that are identical across all columns
spotify_data <- distinct(spotify_data)
```

**Inference:**

Removing identical rows ensures that only unique entries are retained, reducing redundancy in the dataset.

#### 1.6.7 Group Rare Artists into "Other"

```
# Group artists with fewer than the threshold of entries into "Other"
threshold <- 5
artist_counts <- table(spotify_data$artistName)
rare_artists <- names(artist_counts[artist_counts < threshold])
spotify_data$artistName <- ifelse(spotify_data$artistName %in% rare_artists, "Other", spotify_data$arti
```

**Inference:**

Grouping rare artists under the "Other" category reduces the number of unique categories, simplifying the model and making it easier to interpret.

#### 1.6.8 Convert Artist Names to Factor

```
# Convert artistName to a factor
spotify_data$artistName <- as.factor(spotify_data$artistName)
```

**Inference:**

Converting `artistName` to a factor ensures that it is correctly handled as a categorical variable in regression modeling.

---

### 1.6.9 Let's Clean the Genre Column

```r
# Checking the genre data
head(spotify_data$genre)
```

```
## [1] ""                   "british orchestra" "glitchcore"
## [4] "experimental pop"  "album rock"        "alt z"
```

```r
head(unique(spotify_data$genre), 25)
```

```
##  [1] ""                         "british orchestra"
##  [3] "glitchcore"               "experimental pop"
##  [5] "album rock"               "alt z"
##  [7] "guitar case"              "cloud rap"
##  [9] "dance pop"                "desi hip hop"
## [11] "lo-fi sleep"              "contemporary country"
## [13] "bedroom r&b"              "singer-songwriter pop"
## [15] "la pop"                   "lo-fi chill"
## [17] "orchestral soundtrack"    "comic"
## [19] "alternative metal"        "deep underground hip hop"
## [21] "pop"                      "brooklyn drill"
## [23] "classical"                "american orchestra"
## [25] "modern alternative rock"
```

```r
# Count missing values
sum(is.na(spotify_data$genre) | spotify_data$genre == "")
```

```
## [1] 475
```

**Output Interpretation:** This output helps identify how many `genre` values are missing or empty, guiding further cleaning steps.

```r
# Replace empty genre values with "unknown"
spotify_data$genre[spotify_data$genre == ""] <- "unknown"
```

**Inference:** Replacing empty genre values with "unknown" ensures that no missing data is mistakenly treated as an empty string.

```r
# Convert genre names to lowercase
spotify_data$genre <- tolower(spotify_data$genre)
```

```r
# Display the frequency of genres
genre_counts <- sort(table(spotify_data$genre), decreasing = TRUE)
print(genre_counts)
```

```
## 
##                    unknown                      alt z 
##                        475                        328 
##                        pop                      filmi 
##                        301                        206 
##                  dance pop       singer-songwriter pop 
##                         86                         82 
##          alternative metal               anime lo-fi 
##                         75                         68 
##                    art pop                drift phonk 
##                         63                         62 
##                    brostep    modern alternative rock 
##                         58                         56 
##                lo-fi study                        edm 
##                         55                         50 
##                      anime                   chill pop 
##                         49                         47 
##                  classical                      j-pop 
##                         46                         46 
##                  cloud rap                     la pop 
##                         44                         42 
##               modern rock                lo-fi sleep 
##                         40                         37 
##                   boy band                lo-fi chill 
##                         36                         35 
##                baroque pop          german soundtrack 
##                         34                         34 
##           modern indie pop               desi hip hop 
##                         34                         33 
##            icelandic indie                bedroom pop 
##                         33                         32 
##                   desi pop                    hip hop 
##                         32                         32 
##          alternative dance     gen z singer-songwriter 
##                         31                         31 
##                anime score           classic bollywood 
##                         29                         29 
##                   dark r&b                 sad lo-fi 
##                         26                         26 
##              canadian pop                 electropop 
##                         25                         24 
##                   folk-pop                 pov: indie 
##                         24                         24 
##                album rock                  gym phonk 
##                         23                         23 
##                  indie pop           alternative rock 
##                         23                         22 
##                lo-fi beats  canadian contemporary r&b 
##                         22                         21
```

```
##                australian pop                detroit hip hop
##                            20                             20
##                       emo rap                     glitchcore
##                            20                             20
##              indie poptimism                         pop rap
##                            20                             20
##                  bedroom soul                         shanty
##                            19                             19
##                   atl hip hop            canadian electronic
##                            17                             17
##                 post-teen pop                        sad rap
##                            16                             16
##     british singer-songwriter                       hopebeat
##                            15                             15
##                  indietronica                          k-pop
##                            15                             15
##                 alabama indie                     complextro
##                            14                             14
##                       dfw rap                    bedroom r&b
##                            14                             13
##             canadian hip hop                       chillwave
##                            13                             13
##                     downtempo             indie anthem-folk
##                            13                             13
##                 lo-fi jazzhop                      pop dance
##                            13                             13
##                     viral pop                   acoustic pop
##                            13                             12
##                 aesthetic rap                alternative r&b
##                            12                             12
##                   chicago rap           classic pakistani pop
##                            12                             12
##                       traprun                         uk pop
##                            12                             12
##          alternative hip hop                australian dance
##                            11                             11
##                     bass trap              japanese chillhop
##                            11                             11
##                       pop edm                    swedish pop
##                            11                             11
##              video game music               aggressive phonk
##                            11                             10
##                  ambient folk                       bossbeat
##                            10                             10
##                         disco                  electro house
##                            10                             10
##                    glitch hop                      chill r&b
##                            10                              9
##                  classic rock                 easy listening
##                             9                              9
##              french shoegaze                     future bass
##                             9                              9
##         indie game soundtrack                      indie r&b
##                             9                              9
```

```
##                melodic rap                piano rock
##                          9                         9
##                        r&b                afghan pop
##                          9                         8
##            australian rock                   baroque
##                          8                         8
##                   big room               chamber pop
##                          8                         8
##       contemporary country                  hyperpop
##                          8                         8
##              indie pop rap          lgbtq+ hip hop
##                          8                         8
##     modern alternative pop          social media pop
##                          8                         8
##                    sped up               ambient pop
##                          8                         7
##                chill phonk                  chillhop
##                          7                         7
##                  chillstep                   chutney
##                          7                         7
##                  dutch edm               dutch house
##                          7                         7
##                  eurodance                 hollywood
##                          7                         7
##              metropopolis                neo mellow
##                          7                         7
##          new french touch     orchestral soundtrack
##                          7                         7
##         australian hip hop                   chanson
##                          6                         6
##               covertronica                gaming edm
##                          6                         6
##                  gauze pop          indie electropop
##                          6                         6
##            melodic dubstep          modern bollywood
##                          6                         6
##                punjabi pop                 scandipop
##                          6                         6
##       alternative pop rock                 austindie
##                          5                         5
##              brazilian edm        british soundtrack
##                          5                         5
##          conscious hip hop        east coast hip hop
##                          5                         5
##                     grunge         japanese teen pop
##                          5                         5
##          lo-fi brasileiro              new romantic
##                          5                         5
##                    otacore                   pop r&b
##                          5                         5
##            punjabi hip hop                 reggaeton
##                          5                         5
##                      sleep                  speedrun
##                          5                         5
```

```
##                super eurobeat                adult standards
##                             5                              4
##                  afrofuturism                 ambient guitar
##                             4                              4
##               ambient worship                   antiviral pop
##                             4                              4
##             arkansas country                   aussietronica
##                             4                              4
##        australian electropop                         bhangra
##                             4                              4
##               black americana                        bow pop
##                             4                              4
##                  british soul                       broadway
##                             4                              4
##                brooklyn indie                            ccm
##                             4                              4
##           classic country pop                     danish pop
##                             4                              4
##      deep underground hip hop                      denpa-kei
##                             4                              4
##          electro-pop francais                        europop
##                             4                              4
##                   focus beats                     glam metal
##                             4                              4
##                    indie rock                 indonesian r&b
##                             4                              4
##       irish singer-songwriter                     j-acoustic
##                             4                              4
##                         k-rap                     korean r&b
##                             4                              4
##                      la indie                           meme
##                             4                              4
##   nashville singer-songwriter                        nyc pop
##                             4                              4
##                  pakistani pop                   shimmer pop
##                             4                              4
##                    slap house                     soundtrack
##                             4                              4
##                     afrobeats                     anime rock
##                             3                              3
##              australian indie               australian psych
##                             3                              3
##                    bass house                         bhajan
##                             3                              3
##                       boy pop                    chill house
##                             3                              3
##                       country                      dark trap
##                             3                              3
##                        dreamo                      dutch pop
##                             3                              3
##                       electra                    electronica
##                             3                              3
##                           emo                    escape room
##                             3                              3
```

```
##            experimental pop                  filter house
##                           3                             3
##                 garage rock                   hindi indie
##                           3                             3
##                     hip pop                  indian indie
##                           3                             3
##           indie garage rock                 indie rockism
##                           3                             3
##     instrumental post-rock                      irish pop
##                           3                             3
##          modern indie folk                  norwegian pop
##                           3                             3
##                       pixel                      pop folk
##                           3                             3
##               reggae fusion                     scorecore
##                           3                             3
##                  show tunes            slowed and reverb
##                           3                             3
##                 solipsynthm                      teen pop
##                           3                             3
##                 tempe indie                      trip hop
##                           3                             3
##                   viral rap                  alaska indie
##                           3                             2
##                 anime phonk               arab electronic
##                           2                             2
##             australian r&b                  barbadian pop
##                           2                             2
##     bc underground hip hop                 belgian dance
##                           2                             2
##                 belgian edm                 belgian indie
##                           2                             2
##                belgian rock               brazilian house
##                           2                             2
##          british indie rock             british orchestra
##                           2                             2
##             brooklyn drill                 canadian latin
##                           2                             2
##                   candy pop                      cantopop
##                           2                             2
##                        cedm                 chamber psych
##                           2                             2
##                 channel pop                 chicago house
##                           2                             2
##                   chill out  christian alternative rock
##                           2                             2
##          christian hip hop             classic swedish pop
##                           2                             2
##              classic uk pop                         comic
##                           2                             2
##                  dance rock                dancefloor dnb
##                           2                             2
##         deep tropical house                       epicore
##                           2                             2
```

```
##                        eurobeat                       french pop 
##                               2                                2 
##                          g funk            garage rock revival 
##                               2                                2 
##                    german dance                        glam rock 
##                               2                                2 
##                 gregorian dance                        hard rock 
##                               2                                2 
##                           house                       indie folk 
##                               2                                2 
##                  indiecoustica   indonesian singer-songwriter 
##                               2                                2 
##                  industrial pop           israeli mediterranean 
##                               2                                2 
##                     israeli pop                        j-poprock 
##                               2                                2 
##                          j-rock         japanese alternative pop 
##                               2                                2 
##              japanese dance pop                     japanese dub 
##                               2                                2 
##                      jazz cover               korean electropop 
##                               2                                2 
##                      korean pop                    latin hip hop 
##                               2                                2 
##                     leeds indie                      lo-fi cover 
##                               2                                2 
##                     lo-fi indie                         mandopop 
##                               2                                2 
##  melbourne bounce international                melodic metalcore 
##                               2                                2 
##               minneapolis indie                  nashville indie 
##                               2                                2 
##                        neo soul                 new jersey indie 
##                               2                                2 
##              newfoundland indie                        nightcore 
##                               2                                2 
##                 norwegian house                     ohio hip hop 
##                               2                                2 
##                  permanent wave                  pittsburgh rap 
##                               2                                2 
##                           plugg                 stomp and holler 
##                               2                                2 
##                   stutter house               swedish electropop 
##                               2                                2 
##                     swiss indie                    toronto indie 
##                               2                                2 
##                            trap                         uk dance 
##                               2                                2 
##                      uk hip hop                     vapor twitch 
##                               2                                2 
##                       weirdcore                       a cappella 
##                               2                                1 
##                        abstract                   abstract beats 
##                               1                                1 
```

```
##            abstract hip hop                     acoustic opm
##                           1                                1
##                 alabama rap           alternative americana
##                           1                                1
##               ambient lo-fi       american modern classical
##                           1                                1
##          american orchestra                    anime latino
##                           1                                1
##                 anime piano                 ann arbor indie
##                           1                                1
##              anthem worship                       arab trap
##                           1                                1
##           argentine ambient                armenian hip hop
##                           1                                1
##      asian american hip hop                    assamese pop
##                           1                                1
##            atmospheric dnb                      austin rock
##                           1                                1
##  australian alternative pop   australian singer-songwriter
##                           1                                1
##                ballet class                     belgian pop
##                           1                                1
##                bhojpuri pop                           blues
##                           1                                1
##                  blues rock                  brazilian bass
##                           1                                1
##              brighton indie                        britcore
##                           1                                1
##    british alternative rock       british modern classical
##                           1                                1
##             bubblegum dance                        cali rap
##                           1                                1
##        calming instrumental                  canadian celtic
##                           1                                1
##  canadian old school hip hop     canadian singer-songwriter
##                           1                                1
##                 celtic rock                    chennai indie
##                           1                                1
##               chicago indie                  children's folk
##                           1                                1
##                 chill baile  chinese classical performance
##                           1                                1
##          chinese electropop                chinese idol pop
##                           1                                1
##           chinese viral pop                        chiptune
##                           1                                1
##             christian lo-fi                  christian rock
##                           1                                1
##            cincinnati indie               cinematic dubstep
##                           1                                1
##                    city pop                 classic bhangra
##                           1                                1
##   classic progressive house                    classic soul
##                           1                                1
```

```
##          classic texas country          classical performance
##                              1                              1
##                cleveland indie                   colombian pop
##                              1                              1
##          compositional ambient               contemporary r&b
##                              1                              1
##                         cumbia                      cyberpunk
##                              1                              1
##             danish electronic              danish electropop
##                              1                              1
##                  dark clubbing                     dark plugg
##                              1                              1
##                       dark pop                    death metal
##                              1                              1
##                  detroit indie                    disco house
##                              1                              1
##                     diva house                        dmv rap
##                              1                              1
##                    dutch indie                        electro
##                              1                              1
##                   electroclash             english indie rock
##                              1                              1
##                           erhu            estonian electronic
##                              1                              1
##                  ethereal wave                    fingerstyle
##                              1                              1
##       finnish alternative rock                  florida drill
##                              1                              1
##                    folktronica              freeform hardcore
##                              1                              1
##                french hip hop               french indie pop
##                              1                              1
##            french psychedelic                           funk
##                              1                              1
##                   funk carioca                     funk metal
##                              1                              1
##                    future rock                        g-house
##                              1                              1
##                gambian hip hop                      game mood
##                              1                              1
##                   gangster rap                     german pop
##                              1                              1
##                german pop rock                     girl group
##                              1                              1
##                         glitch                    guitar case
##                              1                              1
##                   gujarati pop                       hands up
##                              1                              1
##                      hard bass                   hare krishna
##                              1                              1
##                       hawaiian               hawaiian hip hop
##                              1                              1
##                         hi-nrg                   himachali pop
##                              1                              1
```

```
##                    hmong pop hungarian classical performance
##                            1                               1
##               hypnagogic pop                    indian lo-fi
##                            1                               1
##                    indie emo             indie rock italiano
##                            1                               1
##                   indie soul                      indie viet
##                            1                               1
##         indonesian lo-fi pop             instrumental grime
##                            1                               1
##       instrumental math rock                      irish rock
##                            1                               1
##            italian indie pop           italian library music
##                            1                               1
##                    j-ambient                          j-idol
##                            1                               1
##              j-pop boy group                j-pop girl group
##                            1                               1
##     japanese alternative rock     japanese old school hip hop
##                            1                               1
##               japanese piano                    japanese vgm
##                            1                               1
##                     jazz rap                       josei rap
##                            1                               1
##             k-pop girl group                      kazakh pop
##                            1                               1
##                    latin pop                          lilith
##                            1                               1
##                  liquid funk          lithuanian electronic
##                            1                               1
##                 lo-fi latino                   lo-fi product
##                            1                               1
##                    lo-fi rap                    malaysian pop
##                            1                               1
##               manitoba indie                   melodic drill
##                            1                               1
##                    melodipop                        meme rap
##                            1                               1
##              memphis hip hop               metallic hardcore
##                            1                               1
##           minimal tech house                modern dream pop
##                            1                               1
##             modern folk rock musica indigena latinoamericana
##                            1                               1
## musica portuguesa contemporanea               native american
##                            1                               1
##                neon pop punk          new england americana
##                            1                               1
##                 new wave pop                        nightrun
##                            1                               1
##                    noise pop                    nordic house
##                            1                               1
##                     nu disco         nuevo folklore mexicano
##                            1                               1
```

```
##                        nyc rap                         nz hip hop
##                              1                                  1
##                         nz pop                        odia bhajan
##                              1                                  1
##                       odia pop                      oxford indie
##                              1                                  1
##           pakistani electronic                  pakistani hip hop
##                              1                                  1
##             persian electronic                        persian pop
##                              1                                  1
##               phonk brasileiro                      pinoy hip hop
##                              1                                  1
##                        pluggnb                      pop quebecois
##                              1                                  1
##                       pop soul                        post-grunge
##                              1                                  1
##                       rage rap                           rap rock
##                              1                                  1
##                   romanian pop                      russian drain
##                              1                                  1
##                  russian drill                        russian pop
##                              1                                  1
##                 scottish indie                      shimmer psych
##                              1                                  1
##                singaporean pop   singaporean singer-songwriter
##                              1                                  1
##                    sleaze rock                          soul jazz
##                              1                                  1
##               south african pop       swedish singer-songwriter
##                              1                                  1
##                       swiss pop                        tech house
##                              1                                  1
##                   telugu indie                     thai indie rock
##                              1                                  1
##                     trap queen                        turkish edm
##                              1                                  1
##        uk alternative hip hop            uk contemporary r&b
##                              1                                  1
##             ukrainian viral pop                          vapor pop
##                              1                                  1
##                 virginia indie                       vocal trance
##                              1                                  1
##                           wave                              wonky
##                              1                                  1
```

**Inference:** The frequency table provides insights into the distribution of genres, helping identify the most common and rare genres for possible standardization.

```
# Replace empty strings and unknown genres with NA
spotify_data$genre[spotify_data$genre == "" | spotify_data$genre == "unknown"] <- NA
```

```
# Define a mapping for genre standardization
genre_mapping <- c(
```

```
# General categories
"alt z" = "alternative",
"art pop" = "alternative",
"alternative metal" = "alternative rock",
"modern alternative rock" = "alternative rock",
"alternative pop rock" = "alternative rock",
"alternative r&b" = "r&b",
"alternative dance" = "dance",

# Lo-fi variations
"lo-fi sleep" = "lo-fi",
"lo-fi study" = "lo-fi",
"lo-fi chill" = "lo-fi",
"lo-fi beats" = "lo-fi",
"anime lo-fi" = "lo-fi",
"sad lo-fi" = "lo-fi",

# Pop variations
"post-teen pop" = "pop",
"gen z singer-songwriter" = "pop",
"singer-songwriter pop" = "pop",
"viral pop" = "pop",
"bedroom pop" = "pop",
"pop soul" = "pop",
"alt pop" = "pop",

# Hip hop variations
"atl hip hop" = "hip hop",
"traprun" = "trap",
"cloud rap" = "hip hop",
"emo rap" = "rap",
"sad rap" = "rap",
"brooklyn drill" = "drill",
"drift phonk" = "phonk",
"conscious hip hop" = "hip hop",
"canadian hip hop" = "hip hop",
"ohio hip hop" = "hip hop",

# R&B variations
"dark r&b" = "r&b",
"bedroom r&b" = "r&b",
"chill r&b" = "r&b",
"indie r&b" = "r&b",
"pop r&b" = "r&b",
"uk contemporary r&b" = "r&b",

# EDM and Dance variations
"dance pop" = "edm",
"bass house" = "edm",
"chillwave" = "edm",
"future bass" = "edm",
"edm" = "edm",
"house" = "edm",
```

```
    "eurodance" = "edm",
    "gaming edm" = "edm",

    # Classical variations
    "baroque" = "classical",
    "classical performance" = "classical",
    "german soundtrack" = "classical",
    "orchestral soundtrack" = "classical",
    "baroque pop" = "classical",

    # Rock variations
    "album rock" = "rock",
    "blues rock" = "rock",
    "indie rock" = "rock",
    "modern rock" = "rock",
    "hard rock" = "rock",
    "garage rock" = "rock",
    "grunge" = "rock",
    "alternative rock" = "rock",

    # Indie variations
    "indie pop" = "indie",
    "indie poptimism" = "indie",
    "indie folk" = "indie",
    "indie electropop" = "indie",
    "indie game soundtrack" = "indie",
    "indie anthem-folk" = "indie",
    "indie garage rock" = "indie",

    # Anime and J-Pop
    "anime score" = "anime",
    "anime rock" = "anime",
    "anime phonk" = "anime",
    "j-pop" = "japanese pop",
    "japanese teen pop" = "japanese pop",
    "japanese vgm" = "japanese pop",

    # Other Genre Variations
    "pov: indie" = "indie",
    "ambient guitar" = "ambient",
    "ambient pop" = "ambient",
    "aesthetic rap" = "rap",
    "vapor twitch" = "vaporwave",
    "vapor pop" = "vaporwave",
    "brooklyn indie" = "indie",
    "downtempo" = "ambient",
    "chill phonk" = "phonk",
    "melodic rap" = "rap",
    "neo mellow" = "mellow",
    "post-grunge" = "grunge",
    "emo" = "rock"
)
```

```r
# Standardize genres using the recode function
spotify_data$genre <- recode(spotify_data$genre, !!!genre_mapping)

# Check if the genre column was updated
unique(spotify_data$genre)
```

```
##   [1] NA                               "british orchestra"
##   [3] "glitchcore"                     "experimental pop"
##   [5] "rock"                           "alternative"
##   [7] "guitar case"                    "hip hop"
##   [9] "edm"                            "desi hip hop"
##  [11] "lo-fi"                          "contemporary country"
##  [13] "r&b"                            "pop"
##  [15] "la pop"                         "classical"
##  [17] "comic"                          "alternative rock"
##  [19] "deep underground hip hop"       "drill"
##  [21] "american orchestra"             "scandipop"
##  [23] "alabama indie"                  "stomp and holler"
##  [25] "anime"                          "dfw rap"
##  [27] "punjabi pop"                    "folk-pop"
##  [29] "acoustic pop"                   "filmi"
##  [31] "japanese pop"                   "sleep"
##  [33] "electronica"                    "australian pop"
##  [35] "danish pop"                     "rap"
##  [37] "boy band"                       "solipsynthm"
##  [39] "ambient"                        "show tunes"
##  [41] "hollywood"                      "instrumental post-rock"
##  [43] "canadian contemporary r&b"      "aggressive phonk"
##  [45] "hindi indie"                    "himachali pop"
##  [47] "odia bhajan"                    "desi pop"
##  [49] "classic bollywood"              "afghan pop"
##  [51] "chutney"                        "metropopolis"
##  [53] "shanty"                         "chill pop"
##  [55] "glam rock"                      "brostep"
##  [57] "disco"                          "channel pop"
##  [59] "british singer-songwriter"      "melodic drill"
##  [61] "easy listening"                 "indian lo-fi"
##  [63] "detroit hip hop"                "modern indie pop"
##  [65] "australian dance"               "hip pop"
##  [67] "dutch house"                    "canadian electronic"
##  [69] "video game music"               "k-pop"
##  [71] "irish pop"                      "candy pop"
##  [73] "indie pop rap"                  "neo soul"
##  [75] "classic pakistani pop"          "japanese chillhop"
##  [77] "japanese dub"                   "electra"
##  [79] "social media pop"               "electro house"
##  [81] "bhangra"                        "afrobeats"
##  [83] "gregorian dance"                "electropop"
##  [85] "lo-fi indie"                    "lo-fi jazzhop"
##  [87] "korean r&b"                     "ambient folk"
##  [89] "europop"                        "indie"
##  [91] "dutch edm"                      "indian indie"
##  [93] "pop edm"                        "classic rock"
```

```
##  [95] "glitch hop"                     "canadian pop"
##  [97] "icelandic indie"                "complextro"
##  [99] "piano rock"                     "lo-fi brasileiro"
## [101] "dutch pop"                      "phonk"
## [103] "shimmer pop"                    "italian indie pop"
## [105] "lo-fi latino"                   "bedroom soul"
## [107] "dancefloor dnb"                 "nordic house"
## [109] "pakistani pop"                  "new french touch"
## [111] "pop rap"                        "vaporwave"
## [113] "trap"                           "dark trap"
## [115] "big room"                       "josei rap"
## [117] "super eurobeat"                 "stutter house"
## [119] "australian rock"                "uk dance"
## [121] "minneapolis indie"              "afrofuturism"
## [123] "britcore"                       "french hip hop"
## [125] "leeds indie"                    "minimal tech house"
## [127] "chill out"                      "denpa-kei"
## [129] "brazilian house"                "hands up"
## [131] "bass trap"                      "nu disco"
## [133] "chamber pop"                    "aussietronica"
## [135] "mandopop"                       "indie rock italiano"
## [137] "alaska indie"                   "estonian electronic"
## [139] "bossbeat"                       "romanian pop"
## [141] "antiviral pop"                  "belgian edm"
## [143] "irish rock"                     "modern alternative pop"
## [145] "japanese alternative pop"       "rage rap"
## [147] "nyc pop"                        "hyperpop"
## [149] "instrumental grime"             "russian drain"
## [151] "broadway"                       "singaporean pop"
## [153] "j-poprock"                      "alternative hip hop"
## [155] "british indie rock"             "turkish edm"
## [157] "chillhop"                       "nightcore"
## [159] "classic texas country"          "ukrainian viral pop"
## [161] "abstract beats"                 "indonesian r&b"
## [163] "sped up"                        "chicago rap"
## [165] "instrumental math rock"         "pittsburgh rap"
## [167] "lo-fi cover"                    "colombian pop"
## [169] "french pop"                     "focus beats"
## [171] "gym phonk"                      "nuevo folklore mexicano"
## [173] "hard bass"                      "russian pop"
## [175] "israeli mediterranean"          "israeli pop"
## [177] "thai indie rock"                "j-rock"
## [179] "japanese dance pop"             "chinese viral pop"
## [181] "japanese alternative rock"      "uk pop"
## [183] "indietronica"                   "dance"
## [185] "chillstep"                      "modern indie folk"
## [187] "chanson"                        "bubblegum dance"
## [189] "death metal"                    "vocal trance"
## [191] "mellow"                         "german dance"
## [193] "indie rockism"                  "melodic dubstep"
## [195] "swedish singer-songwriter"      "persian pop"
## [197] "ccm"                            "meme"
## [199] "funk carioca"                   "trap queen"
## [201] "norwegian pop"                  "swiss pop"
```

```
## [203] "slowed and reverb"              "east coast hip hop"
## [205] "meme rap"                       "indiecoustica"
## [207] "chamber psych"                  "g funk"
## [209] "gauze pop"                      "k-rap"
## [211] "new wave pop"                   "finnish alternative rock"
## [213] "calming instrumental"           "nashville indie"
## [215] "australian electropop"          "chill house"
## [217] "pixel"                          "ann arbor indie"
## [219] "bhojpuri pop"                   "reggae fusion"
## [221] "folktronica"                    "teen pop"
## [223] "otacore"                        "native american"
## [225] "alabama rap"                    "belgian dance"
## [227] "latin hip hop"                  "industrial pop"
## [229] "swedish pop"                    "german pop"
## [231] "cleveland indie"                "slap house"
## [233] "jazz cover"                     "wave"
## [235] "funk metal"                     "new england americana"
## [237] "indie viet"                     "bc underground hip hop"
## [239] "virginia indie"                 "chicago house"
## [241] "newfoundland indie"             "austin rock"
## [243] "canadian old school hip hop"    "epicore"
## [245] "british soul"                   "eurobeat"
## [247] "latin pop"                      "australian indie"
## [249] "chiptune"                       "melodic metalcore"
## [251] "viral rap"                      "belgian pop"
## [253] "dreamo"                         "shimmer psych"
## [255] "dutch indie"                    "pop dance"
## [257] "a cappella"                     "garage rock revival"
## [259] "pakistani hip hop"              "covertronica"
## [261] "australian r&b"                 "hopebeat"
## [263] "canadian celtic"                "lo-fi product"
## [265] "british soundtrack"             "italian library music"
## [267] "australian psych"               "new romantic"
## [269] "reggaeton"                      "compositional ambient"
## [271] "pop folk"                       "classic progressive house"
## [273] "bow pop"                        "indonesian singer-songwriter"
## [275] "austindie"                      "indie soul"
## [277] "christian alternative rock"     "adult standards"
## [279] "chinese electropop"             "memphis hip hop"
## [281] "brazilian edm"                  "acoustic opm"
## [283] "armenian hip hop"               "j-idol"
## [285] "children's folk"                "detroit indie"
## [287] "lgbtq+ hip hop"                 "country"
## [289] "celtic rock"                    "scorecore"
## [291] "la indie"                       "electro"
## [293] "anime latino"                   "j-acoustic"
## [295] "swedish electropop"             "canadian singer-songwriter"
## [297] "uk alternative hip hop"         "classic country pop"
## [299] "escape room"                    "speedrun"
## [301] "arab trap"                      "uk hip hop"
## [303] "cantopop"                       "classic swedish pop"
## [305] "dark plugg"                     "glitch"
## [307] "tempe indie"                    "anthem worship"
## [309] "sleaze rock"                    "funk"
```

```
## [311] "arkansas country"                 "chinese idol pop"
## [313] "french shoegaze"                  "swiss indie"
## [315] "weirdcore"                        "liquid funk"
## [317] "odia pop"                         "noise pop"
## [319] "blues"                            "oxford indie"
## [321] "jazz rap"                         "soul jazz"
## [323] "pinoy hip hop"                    "bhajan"
## [325] "gujarati pop"                     "j-ambient"
## [327] "hare krishna"                     "modern bollywood"
## [329] "erhu"                             "fingerstyle"
## [331] "christian lo-fi"                  "musica indigena latinoamericana"
## [333] "disco house"                      "modern folk rock"
## [335] "filter house"                     "hypnagogic pop"
## [337] "german pop rock"                  "atmospheric dnb"
## [339] "dmv rap"                          "irish singer-songwriter"
## [341] "nightrun"                         "korean electropop"
## [343] "arab electronic"                  "dark clubbing"
## [345] "glam metal"                       "korean pop"
## [347] "classic soul"                     "pop quebecois"
## [349] "belgian indie"                    "alternative americana"
## [351] "asian american hip hop"           "phonk brasileiro"
## [353] "nashville singer-songwriter"      "chill baile"
## [355] "hi-nrg"                           "black americana"
## [357] "french psychedelic"               "british alternative rock"
## [359] "japanese piano"                   "russian drill"
## [361] "melbourne bounce international"   "indie emo"
## [363] "cali rap"                         "chennai indie"
## [365] "french indie pop"                 "electro-pop francais"
## [367] "future rock"                      "cinematic dubstep"
## [369] "lithuanian electronic"            "punjabi hip hop"
## [371] "diva house"                       "english indie rock"
## [373] "ambient lo-fi"                    "indonesian lo-fi pop"
## [375] "j-pop boy group"                  "soundtrack"
## [377] "persian electronic"               "christian rock"
## [379] "boy pop"                          "wonky"
## [381] "barbadian pop"                    "girl group"
## [383] "ambient worship"                  "singaporean singer-songwriter"
## [385] "ballet class"                     "florida drill"
## [387] "trip hop"                         "canadian latin"
## [389] "g-house"                          "telugu indie"
## [391] "lilith"                           "rap rock"
## [393] "malaysian pop"                    "brazilian bass"
## [395] "cincinnati indie"                 "nz hip hop"
## [397] "south african pop"                "australian hip hop"
## [399] "neon pop punk"                    "manitoba indie"
## [401] "classic bhangra"                  "abstract"
## [403] "hmong pop"                        "city pop"
## [405] "kazakh pop"                       "cedm"
## [407] "freeform hardcore"                "deep tropical house"
## [409] "contemporary r&b"                 "norwegian house"
## [411] "danish electronic"                "grunge"
## [413] "belgian rock"                     "ethereal wave"
## [415] "cumbia"                           "dark pop"
## [417] "cyberpunk"                        "musica portuguesa contemporanea"
```

```
## [419] "classic uk pop"                      "plugg"
## [421] "anime piano"                         "nz pop"
## [423] "dance rock"                          "chicago indie"
## [425] "modern dream pop"                    "american modern classical"
## [427] "gambian hip hop"                     "tech house"
## [429] "pluggnb"                             "hawaiian hip hop"
## [431] "hungarian classical performance" "argentine ambient"
## [433] "gangster rap"                        "metallic hardcore"
## [435] "british modern classical"            "brighton indie"
## [437] "danish electropop"                   "chinese classical performance"
## [439] "australian singer-songwriter"        "pakistani electronic"
## [441] "christian hip hop"                   "toronto indie"
## [443] "lo-fi rap"                           "japanese old school hip hop"
## [445] "melodipop"                           "new jersey indie"
## [447] "k-pop girl group"                    "scottish indie"
## [449] "permanent wave"                      "abstract hip hop"
## [451] "electroclash"                        "hawaiian"
## [453] "game mood"                           "nyc rap"
## [455] "j-pop girl group"                    "assamese pop"
## [457] "australian alternative pop"
```

**Inference:** This code standardizes the genre column using the mapping defined in `genre_mapping`. The genres are recoded into broader categories, and the final unique values are verified.